INFORMATIQUE THÉORIQUE ET APPLICATIONS

M. MADONIA
S. SALEMI
T. SPORTELLI

On z-submonoids and z-codes

Informatique théorique et applications, tome 25, nº 4 (1991), p. 305-322
<http://www.numdam.org/item?id=ITA_1991__25_4_305_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 25, n° 4, 1991, p. 305 à 322)

ON Z-SUBMONOIDS AND Z-CODES (*)

by M. MADONIA [1], S. SALEMI [1] and T. SPORTELLI [1]

Communicated by J.-E. PIN

Abstract. – *This paper deals with z-submonoids and z-codes. It is shown that the z-submonoid generated by a z-code is free. Moreover, a generalization to the z-codes of the Schützenberger's theorem regarding maximal and complete codes is given: a recognizable z-code is a z-code maximal if it is z-complete.*

Résumé. – *On montre que le z-sousmonoïde engendré par un z-code est libre. En outre, on prouve une généralisation du théorème de Schützenberger sur les codes maximaux et complets : un z-code reconnaissable est un z-code maximal si il est z-complet.*

1. INTRODUCTION

In the framework of automata theory, recent studies [1, 3, 4, 5], have examined the relationship between the languages that are recognized by a two-way automaton and the languages that it is possible to obtain by the closure of a new "zigzag product" on words.

Indeed, in [1], the notions of "zigzag factorization" and "zigzag code" have been introduced and an algorithm to verify if a set of words is a z-code has been given.

In this paper, we have preferred to change the terminology and, for short, the previous terms have been modified in "z-factorization" and "z-code" respectively.

Based on these concepts the paper is organized as follows.

First the point of view is very close to that used in [1].

(*) Received September 1989, revised February 1990.
[1] Università di Palermo, Dipartimento di Matematica ed Applicazioni, via Archirafi, 34, 90123 Palermo, Italy.

In section 2, given a subset X of A^*, we define the set $X^\uparrow$ and we introduce some basic notations.

Afterwards, we define a ternary partial operation in A^*, which we denote by $\uparrow$, and, based on this operation, we define the z-submonoids of A^*, as the subsets of A^* which are stable with respect to $\uparrow$ operation.

Then we show that $X^\uparrow$ is a z-submonoid of A^* and, in particular, that it is the smallest z-submonoid of A^* that contains X.

Moreover we characterize the class of the z-submonoids of A^* and we show that this class is properly included in the class of the submonoids of A^*.

It is also stated that any z-submonoid N of A^* has only one minimal generating system with respect to the $\uparrow$ operation and such a system is denoted by $ZG(N)$. This approach leads to discover that $ZG(N)$ is always included or equal to the minimal generating system of N with respect to the well known $*$ operation.

By using results previously developed in [1], the section 3 deals with the concept of z-code and introduces the definition of trivial z-code.

It is shown that not always $ZG(N)$ is a z-code also when N is a free submonoid of A^*; conversely, it is proved that if $ZG(N)$ is a z-code, then N results also free with respect to $*$ operation.

In the section 4 the definitions of maximal z-code and of z-complete set are given. Using these notions, we obtain a generalization of the well known Shützenberger's theorem regarding maximal and complete codes.

At last, the measure of a z-code is considered in the section 5, and it is shown that there exist some z-complete (or maximal) z-codes which have measure less than 1.

To conclude some open problems are given.

2. DEFINITIONS AND PRELIMINARY RESULTS

Let A be a finite alphabet and A^* the free monoid generated by A. As usual, the elements of A^* are called words and the empty word is denoted by 1. Let $X \subseteq A^*$.

It is possible to define in $A^* \times A^*$ an equivalence relation generated by the set $T = \{((ux, v), (u, xv)) : u, v \in A^*, x \in X\}$.

If $((u, v), (u', v')) \in T$ or $((u', v'), (u, v)) \in T$, then we say that (u, v) *produces in only one step* (u', v'), and we denote this fact by $(u, v) \to (u', v')$.

We call "*step to the right on x*" a step as follows: $(u, xv) \rightarrow (ux, v)$; in the same way $(ux, v) \rightarrow (u, xv)$ is called a "*step to the left on x*". A *path* is a sequence of steps.

With $u \circledR v$ we denote the equivalence class of the pair (u, v).

DEFINITION 1: Given a set $X \subseteq A^*$, $X^\uparrow$ denotes the set:

$$X^\uparrow = \{ w \in A^* : 1 \circledR w = w \circledR 1 \}.$$

This means that a word $w \in A^*$ belongs to $X^\uparrow$ if there exists at least one finite path between the pairs $(1, w)$ and $(w, 1)$. Clearly the first and the last step in the path must be "steps to the right".

The following theorem has been proved in [1]:

THEOREM 1: *For any recognizable $X \subseteq A^*$ there exists an effectively computable deterministic automaton that recognizes $X^\uparrow$.*

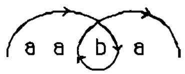
Thus we obtain from the previous theorem that $X^\uparrow \in \mathrm{Rec}(A^*)$ and therefore that $X^\uparrow$ is a rational set.

Example 1: Let $A = \{ a, b \}$ and let $X = \{ a^3 ba^4, a^2 b, b, ba \}$.

The word $w = aaba \notin X^*$ but $w \in X^\uparrow$. Indeed, it suffices to consider the path:

$$(1, w) = (1, aaba) \rightarrow (aab, a) \rightarrow (aa, ba) \rightarrow (aaba, 1) = (w, 1).$$

This path can be visualized as follows:



Remark 1: For any $X \subseteq A^*$ we have $X^* \subseteq X^\uparrow$. In fact, if $w \in X^*$, then $w = x_1 x_2 \ldots x_n$ with $x_i \in X$ for $i = 1, 2, \ldots, n$. Therefore, there exists a path (given by a sequence of steps to the right), as follows:

$$(1, w) = (1, x_1 \ldots x_n) \rightarrow (x_1, x_2 \ldots x_n) \rightarrow \ldots$$
$$\rightarrow (x_1 \ldots x_{n-1}, x_n) \rightarrow (x_1 \ldots x_n, 1) = (w, 1).$$

The converse is not always true, as it has been shown in the example 1.

DEFINITION 2: Given a word $w \in X^\uparrow$, a z-factorization of w over X, of length m, is a sequence of steps $(u_i, v_i) \rightarrow (u_{i+1}, v_{i+1})$ for $i = 1, 2, \ldots, m$ which verifies the following conditions:

1. $u_1 = v_{m+1} = 1$;
2. $v_1 = u_{m+1} = w$;

3. $(u_h, v_h) \neq (u_k, v_k)$ for $h \neq k$.

The condition 3 is necessary to exclude the presence of "cycles" in the z-factorization. In fact, since these cycles should be repeated an arbitrary number of times, they should generate an infinity of different paths from $(1, w)$ to $(w, 1)$, corresponding, indeed, to the same z-factorization of w over X.

DEFINITION 3: Given $w \in X^{\uparrow}$, $l(w, X)$ denotes the minimal length of a z-factorization of w over X.

DEFINITION 4: A z-factorization of $w \in A^*$ is trivial iff its length is equal to 1.

Let us recall the following classical definitions (*see* [2]):

DEFINITION 5: A submonoid of A^* is a subset M which is stable under the concatenation and which contains the neutral element of A^*.

DEFINITION 6: Let M be a submonoid of A^* and let $Y \subseteq A^*$. Y is a minimal generating system of M (with respect to the $*$ operation) if:

- $Y^* = M$
- for any $Z \subseteq A^*$ such that $Z^* = M$ it holds $Y \subseteq Z$.

It is well known that any submonoid M of A^* admits an unique minimal generating system (*see* [2]), which, from now on, we denote by $G(M)$. In particular: $G(M) = (M-1) - (M-1)^2$.

Let us define a new ternary partial *operation* "$\uparrow$" in A^*.

Given $u, v, w \in A^*$ we define:

$$\uparrow(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \begin{cases} \underline{u'v}w' & \text{if } u = u'v \text{ and } w = vw' \text{ with } u', w' \in A^* \\ & \text{undefined otherwise} \end{cases}$$

DEFINITION 7: A z-submonoid of A^* is a subset N which is stable under the $\uparrow$ operation and which contains the neutral element of A^*.

Remark 2: Any z-submonoid of A^* is a submonoid of A^*. In fact it suffices to remark that for any $u, w \in A^*$, $uw = \uparrow(u, 1, w)$. Therefore the $\uparrow$ operation coincides to the concatenation whenever we set $v = 1$.

The converse is not always true: there exist submonoids of A^* that are not z-submonoids of A^*. For example let $M = \{a, aba\}^*$. Of course M is a submonoid of A^*, but it is not a z-submonoid of A^*. In fact if we consider $\uparrow(aba, a, aba) = ababa \notin M$ and thus M is not stable under $\uparrow$ operation.

Remark 3: For any $X \subseteq A^*$, $X^\uparrow$ is trivially a z-submonoid of A^*.

Moreover:

PROPOSITION 1: *For any $X \subseteq A^*$, $X^\uparrow$ is the smallest z-submonoid of A^* that contains X.*

Proof: We have just remarked that $X^\uparrow$ is a z-submonoid of A^* and that $X^* \subseteq X^\uparrow$, so $X \subseteq X^\uparrow$; in order to complete the proof, it suffices to show that, if N is a z-submonoid of A^* that contains X, then $X^\uparrow \subseteq N$.

We set $C_h(X^\uparrow) = \{ w \in X^\uparrow, \text{ such that } l(w, X) = h \}$.

So we have to prove that $C_h(X^\uparrow) \subseteq N$ for every positive integer h. We proceed by induction on h.

For $h=1$ $C_1(X^\uparrow) = X \subseteq N$ and the proposition is trivially true.

Now we suppose that $C_k(X^\uparrow) \subseteq N$ for every $k<h$ and we show that $C_h(X^\uparrow) \subseteq N$.

In fact, let $w \in X^\uparrow$ such that $l(w, X) = h$. Then there exists a z-factorization of w over X of length h, as follows:

$$(1, w) = (1, w_1 w'' w_m) \to (w_1, w'' w_m) \to \dots \to (w_1 w'', w_m) \to (w_1 w'' w_m, 1) = (w, 1)$$

with $w_1, w'', w_m \in A^*$.

We set

$$L_w = \{ x_i \in A^*, \text{ such that the pair } (x_i, y_i) \text{ appears in the } z\text{-factorization of } w \}$$

and

$$R_w = \{ y_i \in A^*, \text{ such that the pair } (x_i, y_i) \text{ appears in the } z\text{-factorization of } w \}.$$

Then let x be the shortest element of L_w that is prefix of w_1 and let y be the shortest element of R_w that is suffix of w_m. With these notations we have:

$$w = \uparrow (x w_i, w_i, w_i y) \quad \text{with} \quad w_i \in A^*,$$

such that $w = x w_i y$ (see *fig.* 1).

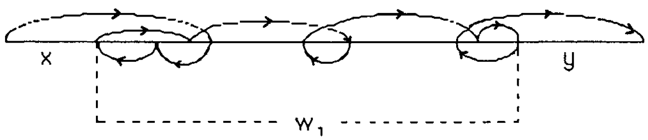


Figure 1

But $w_i \in X^\uparrow$. In fact, in the z-factorization of w over X, there is the subpath

$$\ldots \to (x, w_i y) \to (x_1, y_1) \to \ldots \to (x_t, y_t) \to (xw_i, y) \to \ldots$$

such that:

- $(x, w_i y) \to (x_1, y_1)$ and $(x_t, y_t) \to (xw_i, y)$ are steps to the right
- x is prefix of any x_i for $i=1, \ldots, t$
- y is suffix of any y_i for $i=1, \ldots, t$.

From analogous considerations we have that xw_i, $w_i y \in X^\uparrow$.

Since $l(xw_i, X) < h$, $l(w_i, X) < h$ and $l(w_i y, X) < h$, we have that xw_i, w_i, $w_i y \in N$, by inductive hypothesis. Therefore, since N is stable under the $\uparrow$ operation, $w \in N$ and this completes the proof.

The following proposition 2 characterizes the submonoids of A^* that are also z-submonoids of A^*:

PROPOSITION 2: *Let M be a submonoid of A^* and let $Y = G(M)$. Then M is a z-submonoid of A^* iff $Y^* = Y^\uparrow$.*

Proof: We first show that if $Y^* = Y^\uparrow$, then M is a z-submonoid of A^*.

From $Y = G(M)$ we have $Y^* = M$. But $Y^* = Y^\uparrow$ thus it follows that $M = Y^\uparrow$ and trivially M is a z-submonoid of A^*.

Conversely, let M be a z-submonoid of A^*, $M = Y^*$. Since $Y \subseteq Y^* = M$, we have that M is a z-submonoid of A^* that contains Y. From the proposition 1, we know that $Y^\uparrow$ is the smallest z-submonoid of A^* that contains Y and so $Y^\uparrow \subseteq M = Y^*$. The inclusion $Y^* \subseteq Y^\uparrow$ is trivially true and therefore we have $Y^* = Y^\uparrow$.

Example 2: Let $Y = \{aab, ab, abb, aabb\}$ and let us consider the submonoid of A^*, $M = Y^*$. It is possible to verify that $Y = G(M)$ and that $Y^* = Y^\uparrow$. Therefore M is a z-submonoid of A^*.

Given a z-submonoid N of A^*, let us now define a minimal generating system of N, with respect to the $\uparrow$ operation; from now on, it is called a minimal z-generating system.

DEFINITION 8: Let N be a z-submonoid of A^* and let $X \subseteq A^*$. X is a minimal z-generating system of N if:

- $X^\uparrow = \mathrm{N}$
- for any $Z \subseteq A^*$ such that $Z^\uparrow = N$ it holds $X \subseteq Z$.

Therefore, let X be a subset of A^*; if we consider the z-submonoid $X^\uparrow$ of A^*, not always X is a minimal z-generating system of $X^\uparrow$.

Example 3 Let

$$X=\{a^4, ab, aba^6, aba^3b, aba^3ba^2, aba^2ba, aba^2ba^3, aba^2b^2, aba^2b^2a^2, b, ba^2\}.$$

X isn't a minimal z-generating system of the z-submonoid $X^\uparrow$ of A^*. In fact there exists

$$Z=\{a^4, ab, aba^2ba, aba^2ba^3, b, ba^2\}$$

such that: $Z \subsetneqq X$ and $Z^\uparrow = X^\uparrow$.

The following proposition 3 shows the relationship between a minimal z-generating system of a z-submonoid N and $G(N)$.

PROPOSITION 3: *Let N be a z-submonoid of A^* and suppose that X is a minimal z-generating system of N. Let $Y=G(N)$, it follows that $X \subseteq Y$.*

Proof: Since $Y=G(N)$ and X is a minimal z-generating system of N, we have $Y^* = N = X^\uparrow$. Let $w \in X$. Since $X \subseteq X^\uparrow = Y^*$, w admits a factorization over Y, let it be $w=y_1 \ldots y_n$ with $y_i \in Y$ $i=1, \ldots, n$ and suppose $n>1$. On the other hand, $Y \subseteq Y^* = X^\uparrow$ and, therefore, any word belonging to Y admits a z-factorization over X. This implies that w should admit a non trivial z-factorization over X contradicting the hypothesis that X is a minimal z-generating system. Thus $n=1$ and $w \in Y$.

We now show that any z-submonoid N of A^* has a minimal z-generating system; indeed, we prove that such a system is unique and it is effectively deduced from $G(N)$.

PROPOSITION 4: *Let N be a z-submonoid of A^* and let $Y \subseteq A^*$, $Y=G(N)$. Then the minimal z-generating system of N is unique and it is $(Y-T_Y)$ with $T_Y=\{y \in Y : l(y, Y-y)>1\}$.*

Proof: First we show that $(Y-T_Y)$ is a z-generating system of N, namely that $N=(Y-T_Y)^\uparrow$. First we show that $N \subseteq (Y-T_Y)^\uparrow$. It suffices to verify that any $w \in N$ has a z-factorization over $(Y-T_Y)$. In fact, since $Y=G(N)$ then $Y^*=N$. Thus if $w \in N$ then $w \in Y^*$, *i. e.* $w=y_1 y_2 \ldots y_n$ with $y_i \in Y$, $i=1, \ldots, n$. Suppose that at least one among y_i belongs to T_Y, let it be y_t. Therefore, it should exist a non trivial z-factorization of y_t over Y, *i. e.* it should exist a path:

$$(1, y_t) \rightarrow (y_t', y_t'') \rightarrow \ldots \rightarrow (y_t, 1)$$

with

$$y_t = y_t' y_t'' \qquad \text{and} \qquad y_t', y_t'' \in A^*.$$

Therefore, it is possible to derive the z-factorization of w over $(Y-T_Y)$ as follows:

$$(1, w)=(1, y_1 y_2 \ldots y_n) \rightarrow \ldots \rightarrow (y_1, y_2 \ldots y_n) \rightarrow \ldots$$
$$\rightarrow (y_1 y_2 \ldots y_{t-1}, y_t y_{t+1} \ldots y_n) \rightarrow (y_1 y_2 \ldots y_{t-1} y_t', y_t'' y_{t+1} \ldots y_n) \rightarrow \ldots$$
$$\rightarrow (y_1 y_2 \ldots y_t, y_{t+1} \ldots y_n) \rightarrow \ldots \rightarrow (y_1 y_2 \ldots y_n, 1)=(w, 1).$$

On the other hand $(Y-T_Y)^{\uparrow} \subseteqq N$. In fact $(Y-T_Y) \subseteqq Y \subseteqq Y^*=N$. Therefore N is a z-submonoid that contains $(Y-T_Y)$ and, since $(Y-T_Y)^{\uparrow}$ is the smallest z-submonoid that contains $(Y-T_Y)$, we have that $(Y-T_Y)^{\uparrow} \subseteqq N=(Y-T_Y)^{\uparrow}$.

Now we can prove that $(Y-T_Y)$ is a minimal z-generating system. Suppose that there exists $Z \subseteqq A^*$ such that $Z^{\uparrow}=N$. We show that $(Y-T_Y)$ is contained in Z.

Let $y \in (Y-T_Y)$ then $y \in (Y-T_Y)^{\uparrow}=N=Z^{\uparrow}$; therefore there exists a z-factorization of y over Z. But $Z \subseteqq Z^{\uparrow}=(Y-T_Y)^{\uparrow}$ and this implies that exists also a z-factorization of y over $(Y-T_Y)$. Since $y \notin T_Y$, such a z-factorization has only one step and this step is to the right; it follows that also the z-factorization over Z has only one step and this step is to the right; according to the previous observations it follows that there exists $z \in Z$ such that $y=z$ and $y \in Z$.

From now on, $ZG(N)$ denotes the minimal z-generating system of N, where N is a z-submonoid of A^*.

Remark 4: Given N z-submonoid of A^*, the proposition 4 shows that $ZG(N) \subseteqq G(N)$. This points out that the $\uparrow$ operation is more powerful than the $*$ operation in the class of the z-submonoids of A^*.

Example 4: Let $Y=\{aab, ab, abb, aabb\}$, as in the example 2, and consider $M=Y^*$. We have seen that $G(M)=Y$ and $M=Y^*=Y^{\uparrow}$ is a z-submonoid of A^*. Then it is possible to find the minimal z-generating system of M; in particular $ZG(M)=\{aab, ab, abb\}$. In fact $T_Y=\{aabb\}$, since:

(i) $l(aabb, Y-aabb)>1$; in fact, it suffices to consider the following z-factorization:

$$(1, aabb) \rightarrow (aab, b) \rightarrow (a, abb) \rightarrow (aabb, 1);$$

(ii) any other word of Y belongs to T_Y.

In this case $ZG(M) \nsubseteqq G(M)$.

3. z-CODES AND FREE SUBMONOIDS

An algorithm for testing if a set X is a z-code or not is given in [1]. This test is based on some properties that must be verified by the non-deterministic automaton which recognizes $X^{\uparrow}$.

This section concerns the relationships between z-codes and minimal z-generating systems. Some examples and new results on z-codes and trivial z-codes are presented.

Moreover, it is shown that the minimal z-generating system of a z-submonoid of A^*, free with respect to $*$ operation, is not always a z-code.

Nevertheless, the theorem 3 states that any z-submonoid, which admits as minimal z-generating system a z-code, is free and therefore it has also a minimal generating system that is a code.

DEFINITION 9: A set $X \subseteq A^*$ is a z-code iff any word $w \in A^*$ has at most one z-factorization over X.

Remark 5: If $X \subseteq A^*$ is a z-code, trivially it must be also a code.

Remark 6: If X is prefix or suffix it is easy to see that X is also a z-code; in fact, any word $w \in A^*$ admits at most one z-factorization and this z-factorization is equal to the factorization of w over X. In this case $X^* = X^{\uparrow}$.

Example 5: Let $X = \{a, aba\}$ be a code.

It is easy to see that X is also a z-code. In fact, if we consider the words of A^* which admit a z-factorization with at least one step to the left, they must be as follows:

$$u \quad a \quad b \quad a \quad b \quad a \quad v \qquad \text{with } u, v \in A^*$$

On the other hand, the word $w = ababa$ hasn't any other z-factorization.

Example 6: Let $X = \{a^3 ba^4, a^2 b, ba\}$. X is a code and it is also a z-code. A formal proof that X is a z-code is based on some properties regarding the non-deterministic automaton which recognizes $X^{\uparrow}$ (*see* [1]).

On the other hand, it is not easy to verify, as we have done in the previous example, that X is a z-code, by simple considerations on the words of X.

Example 7: Let $X = \{abb, abba, ba, babb\}$. X is a code, but it isn't a z-code. In fact, the word $w = abbabb$ has two different z-factorizations:

$$(1, abbabb) \to (abb, abb) \to (abbabb, 1)$$
$$(1, abbabb) \to (abba, bb) \to (ab, babb) \to (abbabb, 1).$$

Remark 7: Let X be a z-code. Then $X = ZG(X^{\uparrow})$. In fact, suppose that X isn't the minimal z-generating system of $X^{\uparrow}$; then there exists $Z \subseteqq A^*$ such that $Z^{\uparrow} - X^{\uparrow}$ and $Z \nsubseteq X$. This implies that there exists $x \in X$ such that $x \notin Z$. Since $X \subseteqq X^{\uparrow} = Z^{\uparrow}$, x admits a non trivial z-factorization over Z (this z-factorization is not trivial because $x \notin Z$). But $Z \subseteqq Z^{\uparrow} = X^{\uparrow}$, therefore such a z-factorization over Z gives a non trivial z-factorization of x over X and this is a contradiction being X a z-code.

DEFINITION 10: Let X be a z-code. X is a trivial z-code iff $X^{\uparrow} = X^*$.

Prefix or suffix codes give some examples of trivial z-codes. The code $X = \{a, aabbb, bb\}$, although it is neither prefix nor suffix, is a trivial z-code.

COROLLARY 1: *Let X be a z-code and let $Y = G(X^{\uparrow})$. Then $X \subseteqq Y$. Moreover X is a non trivial z-code iff $X \nsubseteq Y$.*

Proof: It immediately follows from remark 7 and from proposition 3.

In the theory of codes the following theorem is well known (*see* [2]):

THEOREM 2: *If M is a free submonoid of A^*, then $G(M)$ is a code. Conversely if $Y \subseteqq A^*$ is a code, then the submonoid Y^* of A^* is free and Y is its minimal generating system.*

As regards to z-codes, the following problem rises:

PROBLEM: Let $Y \subseteqq A^*$ be a code. By the theorem 2 we have that Y^* is a free submonoid of A^* and $G(Y^*) = Y$. Suppose that Y^* is also a z-submonoid of A^*. By the proposition 4, $ZG(Y^*) = Y - T_Y$. A question obviously rises: such a $ZG(Y^*)$ is always a z-code?

The answer is negative. In fact, it suffices to consider the following example.

Example 8: Let $Y = \{aa, aab, ab, abb, bb\}$. Y is a code then Y^* is free. It is possible to verify that $Y^* = Y^{\uparrow}$ and therefore Y^* is a z-submonoid of A^*. Moreover $Y = ZG(Y^*)$ since $T_Y = \varnothing$. But Y isn't a z-code (for instance, $w = aabb$ is a word which has two distinct z-factorizations over Y).

Nevertheless, the following theorem holds:

THEOREM 3: *Let N be a z-submonoid of A^*. Let $Y = G(N)$ and $X = ZG(N)$. If X is a z-code then Y is a code.*

Proof: Trivially $Y^* = N = X^{\uparrow}$.

In order to prove that Y is a code, it suffices to prove that $u, vw, uv, x \in N$ imply $v \in N$.

Since $Y^*=N=X^\uparrow$, there exist f_1, f_2, f_3 and f_4 z-factorizations over X of u, vw, uv, w respectively.

Let us suppose

$$f_1 : (1, u) \to (u_1, u'_1) \to \ldots \to (u_n, u'_n) \to (u_{n+1}, u'_{n+1}) = (u, 1)$$
$$f_2 : (1, vw) \to (z_1, z'_1) \to \ldots \to (z_r, z'_r) \to (z_{r+1}, z'_{r+1}) = (vw, 1)$$
$$f_3 : (1, uv) \to (t_1, t'_1) \to \ldots \to (t_s, t'_s) \to (t_{s+1}, t'_{s+1}) = (uv, 1)$$
$$f_4 : (1, w) \to (w_1, w'_1) \to \ldots \to (w_m, w'_m) \to (w_{m+1}, w'_{m+1}) = (w, 1)$$

and let us consider the word $uvw \in N$.

If we opportunely combine the z-factorization f_1 with f_2, and f_3 with f_4, we can obtain two z-factorizations over X, f'_1 and f'_2, of the word uvw

$$\begin{aligned} f'_1 : (1, uvw) &\to (u_1, u'_1 vw) \to \ldots \to (u_n, u'_n, vw) \to (u_{n+1}, u'_{n+1} vw) \\ &= (u, vw) \to (uz_1, z'_1) \to \ldots \to (uz_r, z'_r) \to (uz_{r+1}, z'_{r+1}) = (uvw, 1) \\ f'_2 : (1, uvw) &\to (t_1, t'_1 w) \to \ldots \to (t_s, t'_s w) \to (t_{s+1}, t'_{s+1} w) \\ &= (uv, w) \to (uvw_1, w'_1) \to \ldots \to (uvw_m, w'_m) \to (uvw_{m+1}, w'_{m+1}) = (uvw, 1). \end{aligned}$$

Since X is a z-code, f'_1 must be equal to f'_2. Then, suppose $(u, vw) = (t_h, t'_h w)$ with $1 < h < s+1$, and, therefore, $(uz_1, z'_1) = (t_{h+1}, t'_{h+1} w)$.

Let us consider in f'_2 the sequence of steps

$$(t_h, t'_h w) \to (t_{h+1}, t'_{h+1} w) \to \ldots \to (t_s, t'_s w) \to (t_{s+1}, t'_{s+1} w) = (uv, w).$$

We have that u is prefix of t_i and that t_i is a prefix of uv for $i = h, \ldots, s+1$. Thus we can conclude that

$$\begin{aligned} (1, v) = (u^{-1} t_h, v) \to (u^{-1} t_{h+1}, t'_{h+1}) \to \ldots \to (u^{-1} t_s, t'_s) \\ \to (u^{-1} t_{s+1}, t'_{s+1}) = (v, 1) \end{aligned}$$

is a z-factorization of v over X.

Therefore, $v \in X^\uparrow = N$ and the theorem is proved.

4. MAXIMAL Z-CODES AND Z-COMPLETE SETS

The definitions of maximal z-code and of z-complete set are introduced in this section. An interesting result is given in the theorem 5, which establishes the relationship between maximal z-codes and z-complete z-codes. Indeed, this theorem is analogous to the well known Schützenberger's theorem regarding the codes in.

For a more clear exposition, the theorem 5 is preceded by a lemma stating that if X is a z-code such that $G(X^{\uparrow})$ is a maximal code, then X is surely a maximal z-code.

DEFINITION 11: Let $X \subseteq A^*$ be a z-code. X is a maximal z-code over A if it is not properly contained in any other z-code over A. In other words X is a maximal z-code iff $X \subseteq Z$ and Z z-code imply $X = Z$.

DEFINITION 12: Let $X \subseteq A^*$ and $w \in A^*$. The word w is completable in $X^{\uparrow}$ if there exist two words u, $v \in A^*$ such that $uwv \in X^{\uparrow}$.

The set of the words of A^* that are completable in $X^{\uparrow}$ is denoted by $F(X^{\uparrow})$.

DEFINITION 13: Let $X \subseteq A^*$. X is z-complete in A^* if any word $w \in A^*$ is completable in $X^{\uparrow}$.

In other words, X is z-complete in A^* iff $F(X^{\uparrow}) = A^*$.

Remark 8: Let X be a z-complete set and let $Y = G(X^{\uparrow})$. Then Y is complete. In fact, since X is z-complete, $F(X^{\uparrow}) = A^*$. But $X^{\uparrow} = Y^*$, therefore $A^* = F(X^{\uparrow}) = F(Y^*)$ and then the thesis.

LEMMA 1: *Let X be a z-code and let $Y = G(X^{\uparrow})$. If Y is a maximal code, then X is a maximal z-code.*

Proof: Since $Y = G(X^{\uparrow})$, $Y^* = X^{\uparrow}$. Suppose that X isn't a maximal z-code. Therefore there exists $x \in A^*$ such that $x \notin X$ and $X' = X \cup \{x\}$ is a z-code. Note that $x \notin Y$. Indeed, if x should belong to Y, from $Y \subseteq Y^*$, it follows that $x \in Y^* = X^{\uparrow}$; in other words this means that there exists a z-factorization of x over X and such a z-factorization isn't trivial since $x \notin X$. Then x has two distinct z-factorizations over $X \cup \{x\}$ (one is the non trivial z-factorization over X and the other is trivial and it consists of a single step to the right on x) and this is in contradiction with the hypothesis that $X \cup \{x\}$ is a z-code.

Let $N = (X')^{\uparrow}$ be the z-submonoid generated by X'. From the remark 7, we have that $ZG(N) = X'$. Let us show that $Y \cup \{x\} \subseteq G(N)$.

The contradiction will follow: by theorem 3, $G(N)$ is a code and, therefore $Y \cup \{x\}$ is a code which is impossible.

First, $x \in G(N)$ since, from proposition 4, $X' = ZG(N) \subseteq G(N)$. Then let $y \in Y$ and suppose $y \notin G(N)$. Then $y = uv$ where u, $v \in N - 1$. The words u and v have exactly one z-factorization over X' and in one of them a step on x must occur, otherwise $y \notin G(X^{\uparrow}) = G(Y^*) = Y$. On the other hand, as

$y \in Y \subseteq Y^* = X^{\uparrow}$, y has another z-factorization over X' but without steps on x. This is impossible since X' is a z-code. It follows that $Y \subseteq G(N)$ and the lemma has been proved.

Let $Y \in \mathrm{Rec}(A^*)$ and suppose that Y is a code. The following theorem is well know in the theory of codes (*see* [2]):

THEOREM 4: *Y is a complete code iff Y is a maximal code.*

We can prove a theorem analogous to the previous one, holding for the family of the recognizable z-codes:

THEOREM 5: *Let $X \subseteq A^*$ be a recognizable z-code. X is z-complete iff X is a maximal z-code.*

In order to prove the theorem we give a lemma.

LEMMA 2: *Let $X \subseteq A^*$. Suppose that X isn't a z-code and that $w \in A^*$ has two distinct z-factorizations over X. Then, there exists a suffix of w which has two distinct z-factorizations over X, f_1 and f_2, such that the first step of f_1 is different from the first step of f_2.*

Proof: Consider f_1 and f_2 and suppose that the first steps of the two z-factorizations of w are both steps on $x \in X$. We can suppose that there exists, in f_1 or f_2, a step (u, v) such that u is a proper prefix of x.

Let $L_1 = \{ u_i \in A^+$, such that the pair (u_i, v_i) appears in $f_1 \}$ and $L_2 = \{ u'_i \in A^+$, such that the pair (u'_i, v'_i) appears in $f_2 \}$. Then, let u_h be the shortest element of L_1 that is prefix of x and let u'_k be the shortest element of L_2 that is prefix of x. Suppose $|u'_k| \leqq |u_h|$, then v'_k is a suffix of w which has two z-factorizations over X with distint first steps (see *fig.* 2).

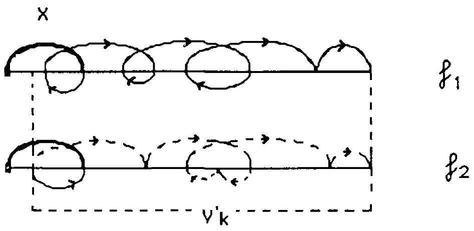


Figure 2

In figure 2, the two distinct z-factorizations of v'_k over X are denoted one by the dotted line and the other one by continuous line.

Proof of the theorem 5. — First we prove that if X is z-complete, then X is a maximal z-code.

Let us consider $X^{\uparrow}$ and let $Y = G(X^{\uparrow})$. From remark 8 it follows that Y is complete and from theorem 3 we know that Y is a code. Moreover, since $X^{\uparrow} \in \mathrm{Rec}(A^*)$, also $Y^* \in \mathrm{Rec}(A^*)$. From previous remarks on Y and from theorem 4 it follows that Y is a maximal code Therefore by lemma 1, X is a maximal z-code.

We now show the converse: if X is a maximal z-code, then X is z-complete.

If $\mathrm{Card}(A) = 1$ this is trivially true. Suppose $\mathrm{Card}(A) > 1$ and suppose that X isn't z-complete. Thus there exists $u \in A^*$ such that $u \notin F(X^{\uparrow})$ Let a be the first letter of the word u and let $b \in A - a$. Let us consider $x = ab^{|u|}$ and $y = ux$. Trivially, $y \notin F(X^{\uparrow})$ [otherwise it should be $u \in F(X^{\uparrow})$ in contradiction with the hypothesis] and y is "unbordered"; this means that any proper prefix of y isn't a suffix of y itself. Moreover, y isn't either prefix, or suffix, or factor of any element of X [otherwise $y \in F(X^{\uparrow})$].

The set $X \cup \{y\}$ is not a z-code since X is a maximal z-code

Then there exists $w \in A^*$ having two distinct z-factorizations, f_1 and f_2, over $X \cup \{y\}$. By the lemma 2, we can choose w such that the first steps of the two z-factorizations are different.

It is useful to remark that:

— both the two z-factorizations must include at least a step on y and this step may be to the left

$$(w' y, w'') \to (w', yw'')$$

or to the right

$$(w', yw'') \to (w' y, w'').$$

In fact, if any of the previous two z-factorizations of w over $X \cup \{y\}$ shouldn't include at least one step on y, then there should exist two distinct z-factorizations of w over X and this leads to a contradiction since X is a z-code. Otherwise, if only one of the two z-factorizations should contain a step on y (doesn't matter if it is to the right or to the left), it should follow $y \in F(X^{\uparrow})$ since $w' yw'' \in X^{\uparrow}$; but this is in contradiction with the fact that y is not completable in $X^{\uparrow}$.

— the occurrences of the factor y in the two distinct z-factorizations can't have "overlap", because y is unbordered. Indeed, if we consider the z-factorizations of w over $X \cup \{y\}$, they contain a step on y and such a step must be to the right: otherwise y should be completable in $X^{\uparrow}$.

From the previous considerations it follows that for any step to the right on y in one of the two z-factorizations of w [for instance, for the step $(w', yw'') \to (w' y, w'')$] there exists, in the same way, a step to the right on y

in the other z-factorization of w [for instance $(v', yv'') \rightarrow (v'y, v'')$ with $v' = w'$ and $v'' = w''$].

In other words, the occurrences of y as a factor in f_1 and f_2 must be "to the right" and "in the same position".

Consider the first occurrences of the factor y in f_1 and f_2: since they must be "to the right" and "in the same position", they don't correspond to the first steps of the two z-factorizations and we have that the step to the right

$$(t_1, yt_2) \rightarrow (t_1 y, t_2) \qquad (*)$$

with $t_1 \in A^+$ and $t_2 \in A^*$, occurs in f_1 and f_2.

Let us take into account the sequence of steps that precede the first step on y in f_1

$$(z_1, z'_1) \rightarrow (z_2, z'_2) \rightarrow \ldots \rightarrow (z_m, z'_m) \rightarrow (t_1, yt_2) \rightarrow (t_1 y, t_2)$$

with $z_i, z'_i \in A^*$ for $i = 1, \ldots, m$ and the sequence of steps that precede the first step on y in f_2

$$(s_1, s'_1) \rightarrow (s_2, s'_2) \rightarrow \ldots \rightarrow (s_r, s'_r) \rightarrow (t_1, yt_2) \rightarrow (t_1 y, t_2)$$

with $s_j, s'_j \in A^*$ for $j = 1, \ldots, r$.

Note that, since $y \notin F(X^\uparrow)$, z_i for $i = 1, \ldots, m$ and s_j for $j = 1, \ldots, r$, are prefix of $t_1 y$.

Let $L_1 = \{z_i \in A^* / 1 \leq i \leq m\}$ and $L_2 = \{s_j \in A^* / 1 \leq j \leq r\}$. Let $z_h \in L_1$ be the element of maximal length in L_1 and let $s_k \in L_2$ be the element of maximal length in L_2. Suppose $|z_h| \geq |s_k|$. Then $z_h \in X^\uparrow$ and it has two distinct z-factorizations over X derived by a suitable combination of steps of f_1 and f_2 (see *fig.* 3).

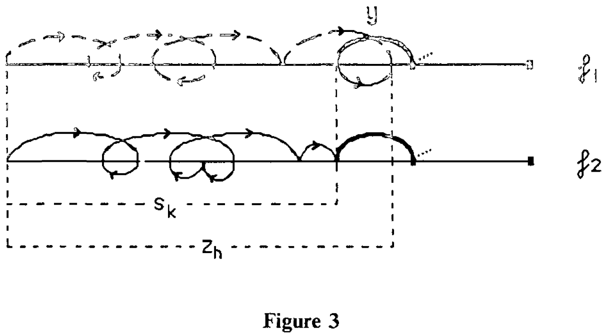


Figure 3

In figure 3, the two distinct z-factorizations of z_h over X are denoted one by the dotted line and the other one by the continuous line.

But this is in contradiction with the hypothesis that X is a z-code and the theorem is proved.

Remark 9: Note that, in the theorem 5, to show that if X is a maximal z-code then X is complete, the assumption that X is recognizable isn't necessary, but this assumption is essential to show the converse.

Remark 10: Let $X \subseteq A^*$ be a z-code and let $Y = G(X^{\uparrow})$. We have just seen (lemma 1) that if Y is a maximal code then X is a maximal z-code. The converse follows from the theorem 5. Indeed, if X is a maximal z-code then X is z-complete and therefore, from the remark 8, Y is a complete code. From the theorem 4, it follows that Y is a maximal code.

5. SOME PROPERTIES OF THE MEASURE OF A Z-CODE

Let A be a finite alphabet with cardinality $|A|$ and let $X \subseteq A^*$ be a code. It is well known that the inequality of Kraft-Mcmillan holds:

$$\alpha(X) = \sum_{x \in X} |A|^{-|x|} \leqq 1.$$

If X is finite with cardinality $|X| = n$, the previous series becomes a finite sum of n terms.

The value $\alpha(X)$ is called measure of the set X.

Trivially if $X \subseteq Y$ then $\alpha(X) \leqq \alpha(Y)$ [if $X \subsetneqq Y$ then $\alpha(X) < \alpha(Y)$].

In the theory of codes it is known that the inequality of Kraft-Mcmillan gives a simple method for testing whether a code is maximal and then complete; in fact, let X be a code; then $\alpha(X) = 1$ if and only if X is maximal (*see* [2]).

Remark 11: Trivially the inequality of Kraft-Mcmillan holds also if X is a z-code. Moreover, if X is a non trivial z-code and $Y = G(X^{\uparrow})$, then Y is a code and $X \subsetneqq Y$; it follows that a non trivial z-code has always measure < 1.

Remark 12: If X is a non trivial z code, then $\alpha(X) < 1$ and this inequality holds also for X maximal z-code and therefore for X z-complete. It follows that, for a non trivial z-code X, it is not possible to decide whether it is z-complete or not with a simple check on the value of its measure.

Example 9: Let $X=\{a^2, ab, ab^2, b^3, ba^3, ba^2b, baba, bab^3\}$. X is a code. The inequality $\alpha(X)<1$ holds, then X is not a complete code in A^*, but it is completable. It suffices to add the word $w=ba^2b^2$.

X is also z-code and, since $w \in X^\dagger$, X is z-complete.

It follows that X is a z-complete z-code and its measure is <1.

SOME OPEN PROBLEMS

PROBLEM 1 (Chap. 2) In the proposition 3 it is stated that, for any z-submonoid N of A^*, $ZG(N) \subseteqq G(N)$. It is easy to see that there exist z-submonoids N of A^* such that $ZG(N)$ is finite, although $G(N)$ is an infinite set.

Example: Let $N=X^\dagger$ with $X=\{a, aba\}$. Then

$$ZG(N)=X \qquad \text{and} \qquad G(N)=\{a(ba)^*\}.$$

Characterize the z-submonoids N such that $ZG(N)$ is finite and $G(N)$ is infinite.

PROBLEM 2 (Chap. 3) :Referring to the definition of trivial z-code, we have shown that there exist trivial z-codes which are neither prefix, nor suffix. Characterize the family of trivial z-codes.

PROBLEM 3 (Chap. 3). — Let N be a z-submonoid of A^*, that is free with respect to $*$ operation. We have remarked that $ZG(N)$ is not always a z-code (*see* example 8).

Characterize those z-submonoids N of A^* that are free with respect to $*$ operation and such that $ZG(N)$ results a z-code.

PROBLEM 4 (Chap. 5) : In the theory of codes it is known that any complete set X has measure $\alpha(X) \geqq 1$. This property does not hold for z-complete sets (*see* example 9).

In the interval [0, 1] find, if it exists, a lower bound for the measure of a z-complete set.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous referees for their helpful recommendations and suggestions.

REFERENCES

1. M. ANSELMO, Automates et codes zigzag, *R.A.I.R.O. Inform. Théor. Appl.*, 1991, *25*, 1, pp. 49-66.
2. J. BERSTEL and D. PERRIN, Theory of codes, *Academic Press*, 1985.
3. J. C. BIRGET, Two-way automaton computations, *R.A.I.R.O. Inform. Théor. Appl.*, 1990, *24*, 1, pp. 47-66.
4. J. P. PÉCUCHET, Automates boustrophédons, langages reconnaissables de mots infinis et variétés de semi-groupes, *Thèse d'État*, L.I.T.P., mai 1986.
5. J. P. PÉCUCHET, Automates boustrophédons, semi-groupe de Birget et monoïde inversif libre, *R.A.I.R.O. Inform. Théor. Appl.*, 1985, *19*, 1, pp. 71-100.